FRVT 1:N Remaining Errors in Cooperative FR: Implications for Quality Assessment

patrick.grother@nist.gov

IFPC 2018 @ NIST

November 29, 2018

IFPC 2018 - Tuesday Nov 27

		0800 Arrive + Registration: 08:30 Welcome + Logistics
Government and Borders	01	0840 **Arun Vemury**, DHS Science + Technology Directorate: *Welcome + DHS context*
	02	0850 **Dan Tanciar,** US Customs and Border Protection: *CBP use of Facial Recognition in Development of a Biometric Entry/Exit System*
	03	0915 **Oliver Bausinger**, BSI, *Smart Borders: EES, ETIAS and Interoperability - towards a unified identity management for Third Country Nationals*
	04	0940 **Anna Stratmann**, BSI: *Biometric processes of the Entry Exit System*
		1005 Break
	05	1035 **Markus Nuppeney**, BSI: Automated Border Control (EasyPASS): *Monitoring the system performance*
	06	1100 **James L. Wayman, John P. Bowes and Joshua Abraham**, 2018 for Department of Home Affairs, Australia, *SmartGate(TM) Update*
	07	
Fast Cap	08	1150 **John Howard**, SAIC: *Evaluation of rapid face capture devices*
	09	1215 **Ilan Arnon**, Face4Systems: *Face recognition on-the-move: Case Studies*
		1240 Lunch
FR in Police	10	1340 **Geoff Whitaker**, DSTL UK: *ISO 30137 video surveillance and OSAC ASTM update E3115*
	11	1405 **Mark Branchflower**, Interpol: *Face recognition in Transnational Crime*
	12	
		1430 Break (cafeteria closes at 1500)
Demographics	13	1500 **John Campbell**, Bion Biometrics: ISO/IEC 22116 *Differential impacts of demographics in biometric systems*
	14	1520 **Yevgeniy Sirotin,** SAIC: *Estimating relative skin reflectance and measuring its effect on recognition.*
	15	1545 **Mike King**, Florida Institute of Technology: *Demographic effects in face recognition*
	16	1610 **Clare Garvie**, Center on Privacy & Technology Georgetown Uni.: *Consequences of differential impacts*
	17	1635 **Patrick Grother**, NIST: *Demographic dependencies in contemporary face recognition algorithms*
	18	1700 **Panel on Demographics: John Campbell, Clare Garvie, Patrick Grother, Mike King, Yevgeniy Sirotin**
		Talks: 16. Dress code: Business casual, face masks

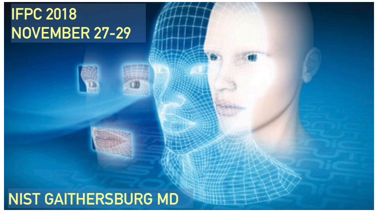
IFPC 2018 - Wednesday Nov 28

		0800 Arrive + Registration 08:30: Start + Review Day 1
Security	19	0835 **Lars Ericson**, IARPA: *Overview of the Odin program on presentation attack detection*
	20	0855 **Ralph Breithaupt**, BSI: *Presentation Attack Detection & Morphing: New developments in biometric security testing and certification*
	21	0920 **Rasa Karbauskaite**, Frontex: *Morphing and other related vulnerabilities for border control*
	22	0940 **Max Dermann**, Bank of New Zealand: *Evaluation of face PAD solutions - a bank's journey*
Morphing	23	1005 **Gert Jan de Nijs**, Dutch Vehicle Authority: *Creating a process to prevent photo fraud*
		1030 Break
	24	1055 **Fons Knopjes**, Passports Netherlands: *SOTAMD: A European state of the art morph detection program*
	25	1120 **Kari Kanto**, The National Police Board of Finland: *Morph detection experiments with large data sets*
	26	1145 **Christoph Busch**, HDA/NTNU: *Morphing attack detection overview*
	27	1210 **Marta Gomez-Barrero,** Hochschule Darmstadt: *Vulnerability Evaluation of Presentation + Morphing Attacks*
	28	1230 **Mei Ngan,** NIST: *FRVT Face Morph Detection Evaluation*
		1245 Lunch
Human Aspects	29	1400 **Jonathon Phillips**, NIST: *Recognition Accuracy of Forensic Examiners, Super-recognizers, and Algorithms*
	30	1420 **Richard Vorder Bruegge**, FBI: *Improving the Process: What could help forensic examiners make better decisions?*
	31	1445 **Eilidh Noyes,** University of Huddersfield: *What is a super-recognizer?*
		1510 Break (cafeteria closes at 1500)
	32	1540 **Carina Hahn**: NIST: *Issues on measuring facial forensic apprenticeship*
	33	1605 **David White**, UNSW-Sydney: *Incorporating human perceptual expertise in face identification systems*
	34	1630 **Carlos Castillo**, Uni. of Maryland: *DCNNs for unconstrained face recognition*
	35	1655 **Alice O'Toole**, UT Dallas: *Understanding face representations in deep convolutional neural networks: Face Space Theory evolves*
	36	1720 **Neal Gieselman**, Aware Inc.: *Tools for human face comparison*
	37	
		Talks: 18 Social Event 6PM: Dogfish Ale House opposite NIST

IFPC 2018 - Thursday Nov 29

		0800 Arrive + Registration 08:30: Start + Review Day 2
Passports	38	0840 **Amir Arien**, Biometrics Registration Authority, Israel: *The Israeli National Biometric Project*
	39	0905 **Delia McGarry and Stephen Melsom**, U.S. Department of State: *Image manipulation detection and effects of perspective distortion on face identification*
	40	0930 **Shashi Samprathi**, Australian Passport Office DFAT: *Update on uses of face recognition*
	41	0955 **Andreas Wolf**, Bundesdruckerei: *ICAO's technical report on portrait quality*
		1020 Break
Privacy	42	1050 **Mickey Cohen**, Shanit: *Privacy, Security and facial de-identification aspects*
	43	1110 **Arun Ross,** Michigan State University: *Semi Adversarial Networks for Face De-identification*
Testing, Standards, Benchmarks	44	1135 **Stephane Gentric**, Idemia: *TBA*
	45	1200 **Thorsten Thies**, Cognitec: *Effects of wrong ID labels*
	46	1225 **Brendan Klare**, Rank One Computing: *Emerging applications in commercial face recognition*
		1240 Lunch
	47	1340 **Christoph Busch**, Hochschule Darmstadt: *Measures for benchmarking indexing algorithms*
	48	1405 **Michael Thieme**, Novetta: *Impact of Non-Facial Regions on FR Performance*
	49	1430 **Tony Mansfield**, NPL: *ISO/IEC 30137-2 Biometric video surveillance - testing and reporting*
		1450 Break (cafeteria closes at 1500)
	50	1520 **Marek Rejman-Greene**, IdentityForServices: *Design and management of reliable services using face recognition*
	51	1540 **Chris Malec**, DSTO: *Australian government FR algorithm performance testing*
	52	1605 **Matt Pruitt**, NEC: *Getting the Best Facial Image in an Uncontrolled Environment: The Effect of User Experience on Facial Quality and Match Scores*
	53	1630 **Nathan Kalka + Brianna Maze,** Noblis: *Curating large-scale face recognition benchmark test sets*
	54	1700 **Patrick Grother**, NIST: *FRVT 2018 and the errors that remain in FR systems: Future Image Quality Standardization*
		Talks: 17 Adjourn: Until 2020

NIST



Homeland Security
Science and Technology

NPL
National Physical Laboratory

European Association for Biometrics
eab
Human Identity in Europe

FRVT 1:N 2018:

The largest public independent face recognition test ever conducted

Phase 1+2 report availability

NISTIR 8238 Draft

Ongoing Face Recognition Vendor Test (FRVT)
Part 2: Identification

Patrick Grother
Mei Ngan
Kayee Hanaoka
Information Access Division
Information Technology Laboratory

This document is circulated as a draft for comment.

2018/05/22

National Institute of Standards and Technology
U.S. Department of Commerce

Interagency Report 8238
- Published 2018-11-26

Attributes accuracy with vendor names.

Revision follows 2019-01
- Add result for Phase 3 algorithms 2018-10-30
- Compare enrollments styles
  - Consolidated
  - Unconsolidated
- Add ageing
- Add selectivity metric
- Delete extraneous images
- Larger database

Participation in Phase 1+2 (June 2018)

45 Developers of 127 Algorithms from 13 countries

- 3divi (RU)
- Alchera (KR)
- Aware (US)+
- Ayonix (JP)
- Camvi (US)
- Cogent Gemalto (FR)+
- Cognitec (DE)
- Dermalog (DE)+
- Ever AI (US)
- Eyedea (CZ)
- Glory (JP)
- Gorilla (CN)
- HB Innovation (KR)
- HIK Vision (CN)

- Idemia (FR)+
- Imagus (AU)
- Incode (US)
- Innovatrics (SL)+
- Innovation Sys. (RU)
- Megvii / Face++ (CN)
- Microfocus (US)
- Microsoft (US)+
- NEC (JP)+
- Neurotechnology (LI)+
- NTechLab (RU)
- Rank One (US)
- Real Networks (US)

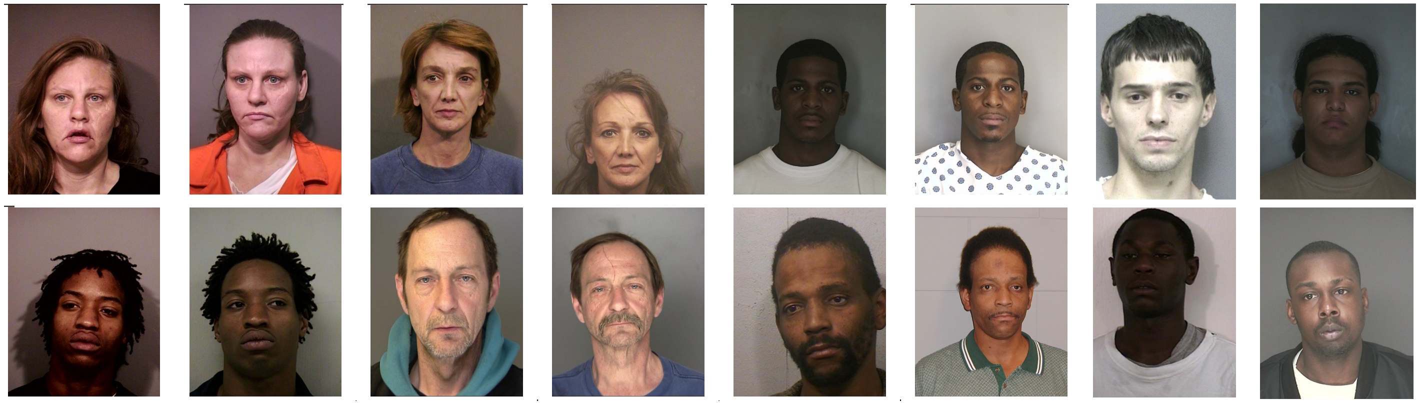
- Shaman (US)
- SIAT CASIA (CN)
- Smilart (RU)
- Synesis (RU)
- Tevian (RU)
- Tiger IT (BG)+
- Tong Yi Trans (CN)
- Vigilant Solutions (US)
- Visidon (FI)
- Visionlabs (RU)
- Vocord (RU)
- Yisheng (CN)
- Yitu (CN)

\+ = Multimodal

Not participating:

- Amazon
- Google
- Facebook
- IBM
- Element AI
- ... many others

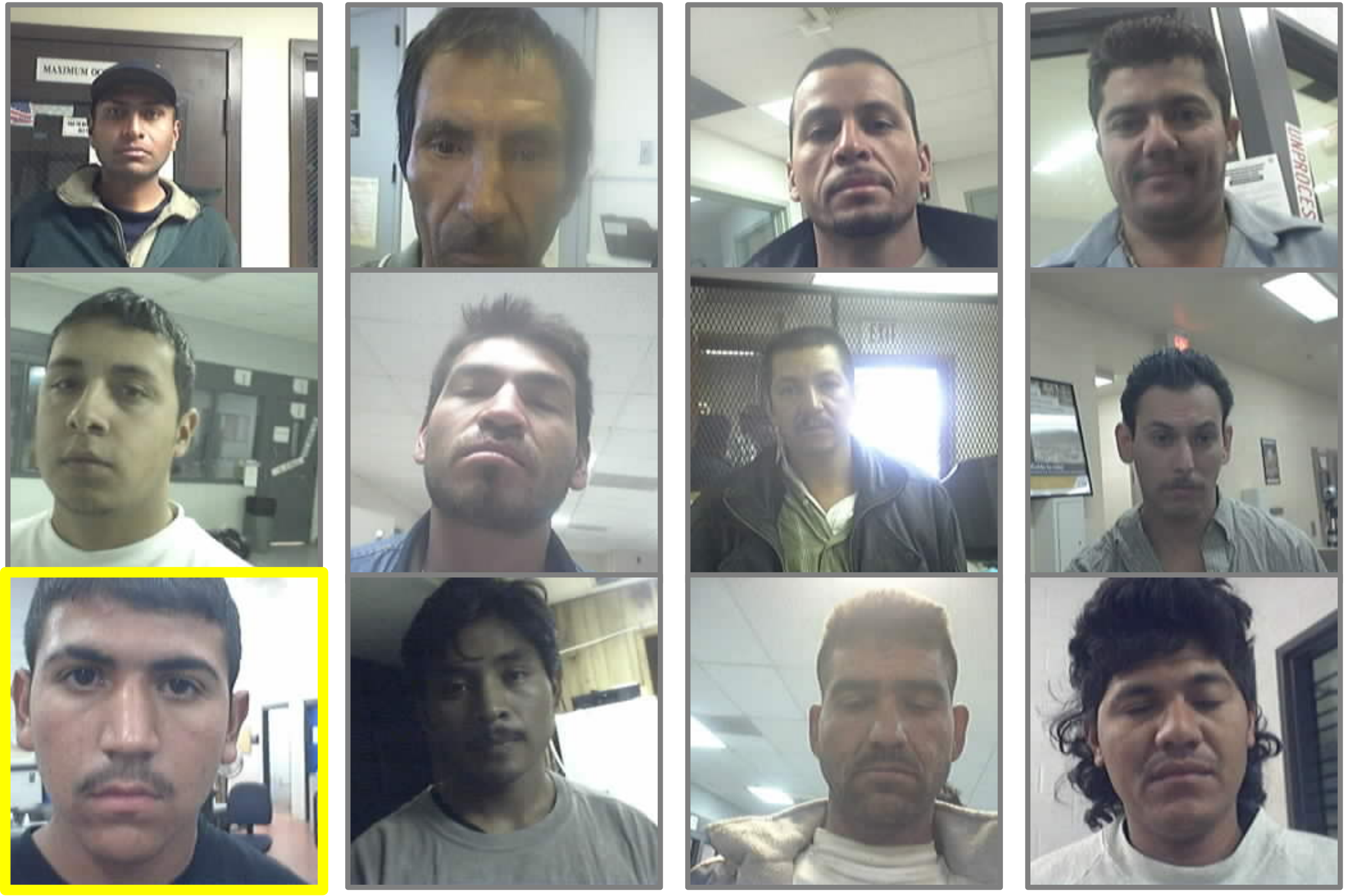
FRVT 2018 Mugshots / Booking / Charge Photos



Images excluded:

1. Webcams (240x240, two shown)
2. Profile views, and others where intent was something other than frontal
3. Tattoos and other non-face images

Operational Webcam images (these from MEDS DB)



NIST

Mugshot-to-mugshot	
Threshold set for FPIR = 0.01	N = 1.6 million people 1.6 million images
Yitu	0.011
Microsoft	0.013
SIAT	0.009
Visionlabs	0.022
Ever AI	0.023
NTechLab	0.024
Idemia	0.024
Neurotechnology	0.030
I-Systems	0.035
Cogent	0.032
NEC	0.049
Cognitec	0.055
HIK Vision	0.056
Megvii / Face++	0.058
RankOne	0.073

FNIR x 2.6

FNIR x 4.1

Webcam-to-mugshot	
Threshold set for FPIR = 0.01	N = 1.6 million people 1.6 million images
Yitu	0.028
Microsoft	0.053
SIAT	0.46 (!!)
NTechLab	0.065
Megvii / Face++	0.067
Neurotechnology	0.073
Ever AI	0.074
Idemia	0.079
I-Systems	0.080
Visionlabs	0.087
NEC	0.093
Cogent	0.100
HIK Vision	0.101
Cognitec	0.135
RankOne	0.187

Developers of Most Accurate Algorithms:

Image Quality Matters

Explaining residual errors from Microsoft and Yitu

Of the 600 mates not returned on candidate lists of length 50

- ~30% Body tattoo containing a face
- ~15% Different person (truth labelling error)
- Profile views
  - 40% for Yitu
  - 30% for Microsoft
- Poor quality (scanned, dim, low contrast)
  - 15% for Yitu
  - 25% for Microsoft

Estimates from 600 false negatives, produced in M = 154549 searches with FNIR(N, R, T) = 0.004

- N = 12,000,000
- R = 50
- T = 0

High-scoring non-mates
Images above T for FPIT(T) = 0.001

- ~20% Confident different person (doppelganger)
- Same person different ID
  - ~44% Confident
  - ~33% Not confident, twin?
  - ~2% Confident, same photo session
- ~1% Scanned image, low quality

Estimates from 660 images involved in 330 false positives, produced in M = 331254 searches with FPIR(N, T) = 0.001

- N = 12,000,000

New FRVT Result 2018-11-26: Profile searches

N = 1.6M	Rank 1 Hit Rate	Rank 50 Hit Rate
ALG-A	91%	94%
ALG-B	73%	85%
ALG-C	17%	21%
ALG-D	87%	93%

Face quality drivers

Background

- » Increased reliance on face recognition
- » Increased use globally, with interchange
- » Unlike fingerprint, iris, most face cameras are "dumb", unware of the face itself
- » Many photos deviate from ISO/ICAO
  - Subject appearance
  - Poor imaging
- » Better recognition algorithms
  - But fail with pose, resolution, demographics
- » Human "forensic" adjudication errors
- » New opportunities for image manipulation

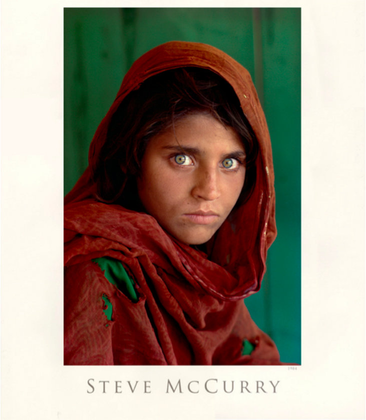
Short terms solutions

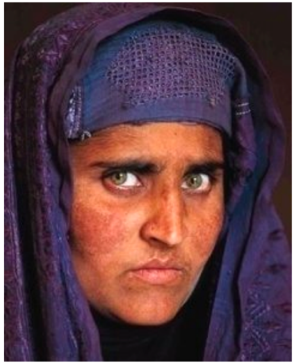
- » Better face recognition algorithms
- » Quality assessment
  - At capture time
  - Over an enterprise
  - Imaging systems

Longer term solutions

- Face-aware capture devices

The March of Time: Ageing





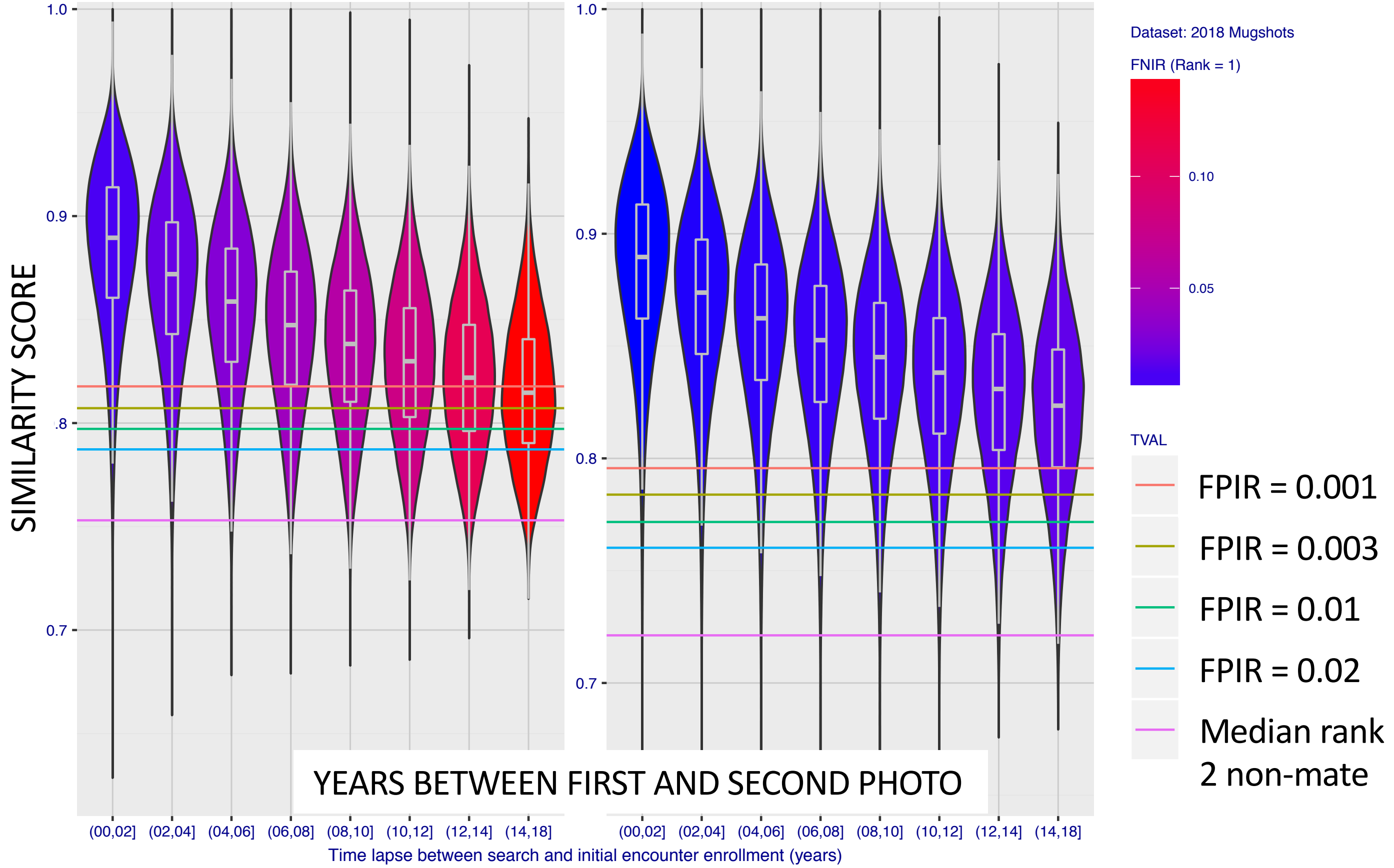

c. National Geographic, photographic portrait by journalist Steve McCurry, 1984

Ageing

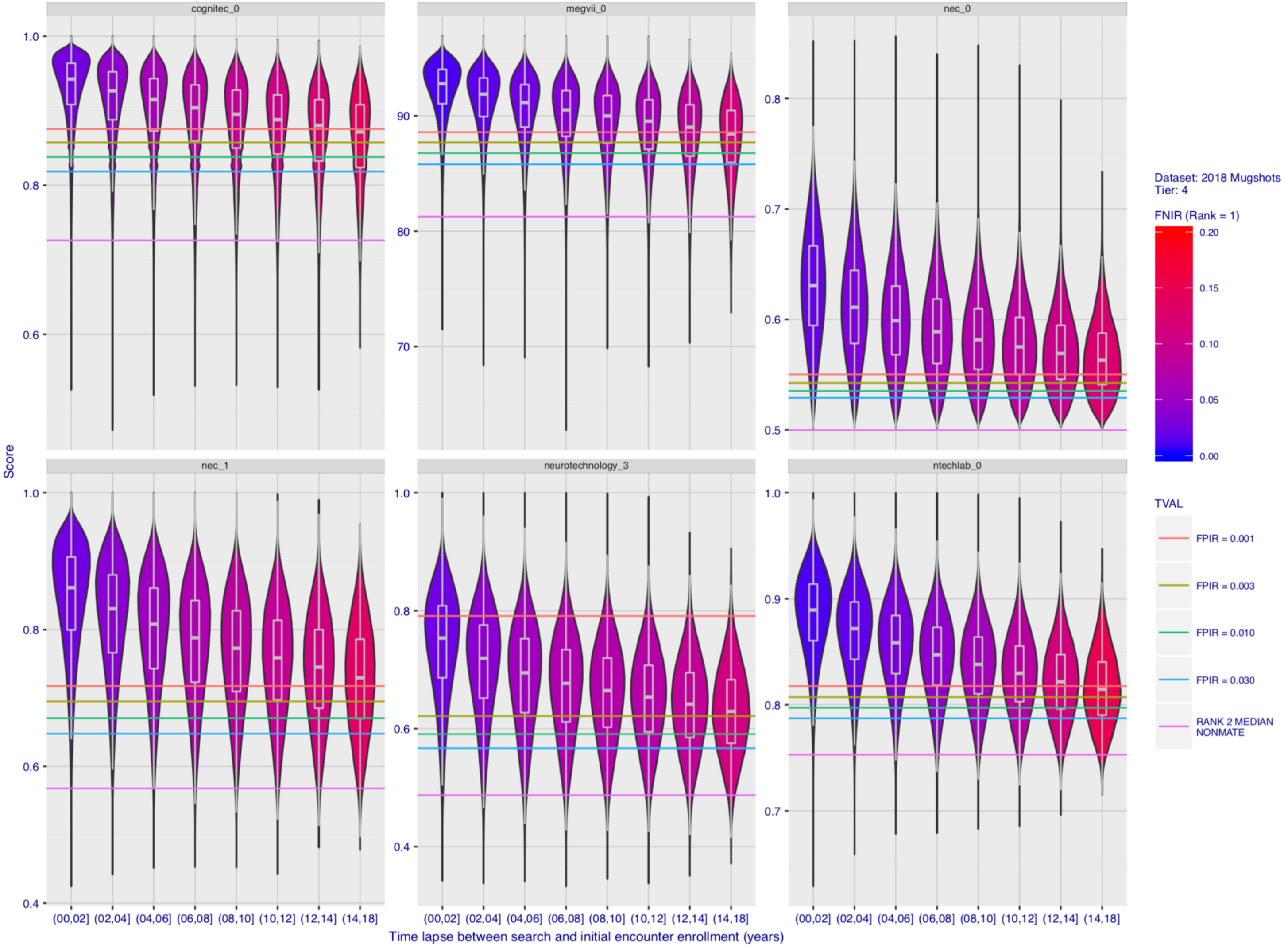
2002-08 2004-10 2010-05 2012 2013-08 2018-06

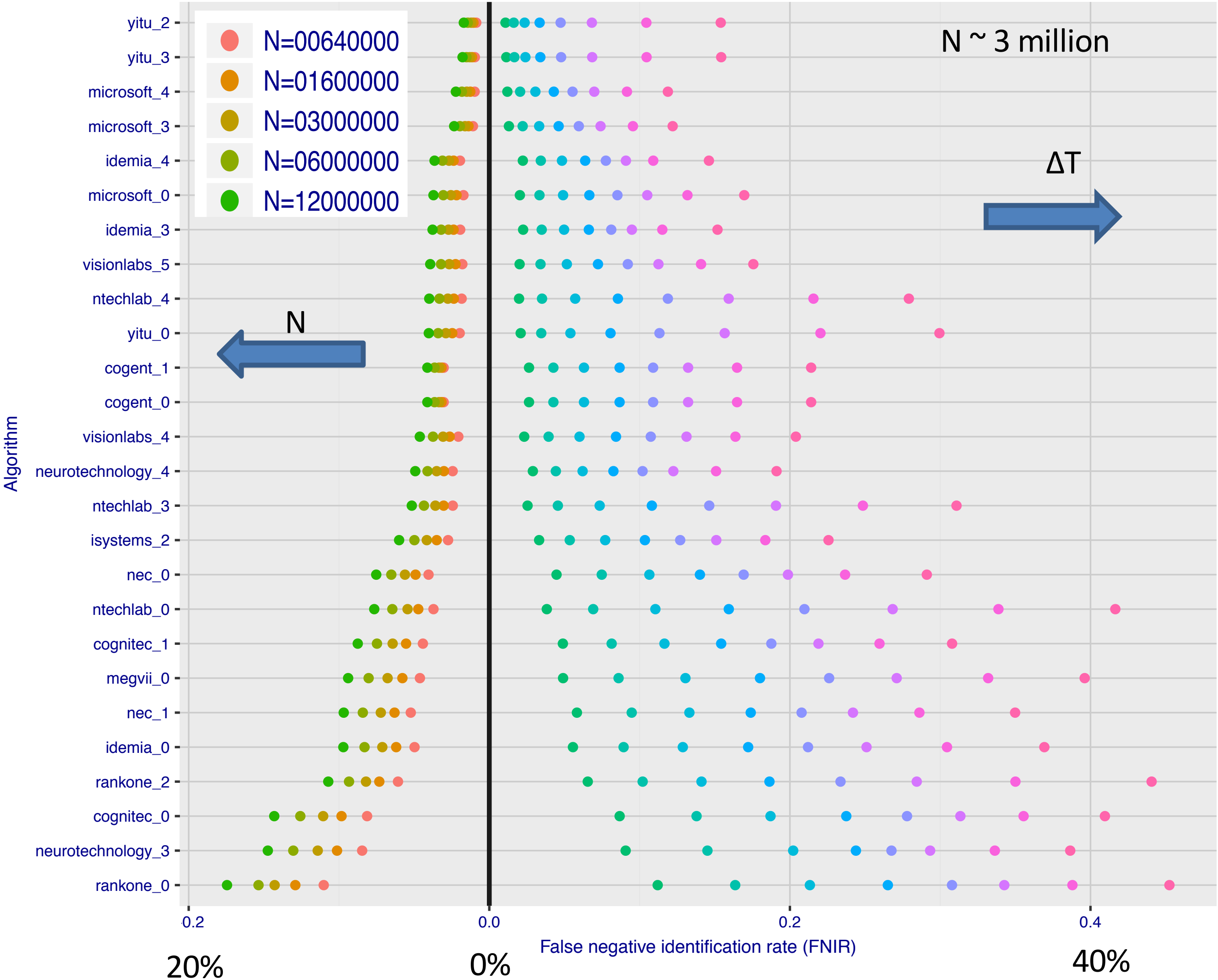
https://www.bellingcat.com/news/uk-and-europe/2018/09/26/skripal-suspect-boshirov-identified-gru-colonel-anatoliy-chepiga/

Mate score distributions under ageing



Ageing: Further Algorithms





NIST

Performance in perspective: What matters more?

1. Algorithm
2. Population size
3. Ageing

Children: Age and Ageing

FNMR — **Age Variation (Years)** — TIME-LAPSE

AGE — **Youngest Age of Child In Image Pair (Years)**

	0	1	2	3	4	5	6	7	8	9	10
0	0.344	0.477	0.568	0.652	0.704	0.731	0.783	0.837	0.897	0.933	0.951
1	0.052	0.119	0.166	0.248	0.287	0.335	0.418	0.497	0.590	0.698	0.768
2	0.034	0.041	0.100	0.134	0.183	0.205	0.266	0.324	0.417	0.553	0.662
3	0.050	0.053	0.073	0.106	0.124	0.155	0.200	0.266	0.361	0.505	0.615
4	0.028	0.051	0.065	0.082	0.096	0.118	0.168	0.237	0.360	0.505	0.599
5	0.033	0.048	0.049	0.069	0.076	0.101	0.149	0.224	0.361	0.476	0.561
6	0.029	0.031	0.033	0.048	0.069	0.092	0.151	0.232	0.360	0.453	0.525
7	0.032	0.023	0.020	0.046	0.065	0.093	0.160	0.249	0.337	0.412	0.456
8	0.024	0.013	0.033	0.046	0.070	0.105	0.174	0.242	0.312	0.369	0.422
9	0.008	0.011	0.038	0.054	0.084	0.118	0.172	0.227	0.273	0.328	0.370
10	0.015	0.012	0.028	0.052	0.089	0.119	0.166	0.199	0.240	0.284	0.325
11	0.005	0.014	0.024	0.076	0.091	0.116	0.143	0.187	0.219	0.260	0.286
12	0.025	0.020	0.037	0.056	0.076	0.093	0.121	0.151	0.187	0.217	0.254
13	0.008	0.021	0.038	0.051	0.061	0.070	0.098	0.116	0.141	0.167	0.202
14	0.011	0.012	0.020	0.036	0.043	0.055	0.070	0.085	0.106	0.126	0.152
15	0.005	0.011	0.013	0.026	0.034	0.043	0.056	0.069	0.083	0.101	0.129
16	0.008	0.010	0.018	0.024	0.027	0.034	0.043	0.053	0.068	0.089	0.095
17	0.009	0.006	0.011	0.018	0.023	0.028	0.037	0.043	0.057	0.069	0.088

Worst Performance | Midpoint | Best Performance

D. Michalski, S.Y. Yiu, C. Malec, *The Impact of Age and Threshold Variation on Facial Recognition Algorithm Performance using Images of Children*, ICB 2018, Surfers Paradise.

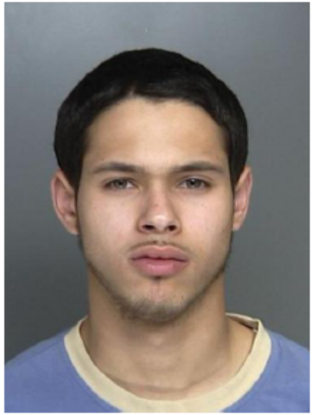
Face Image Quality

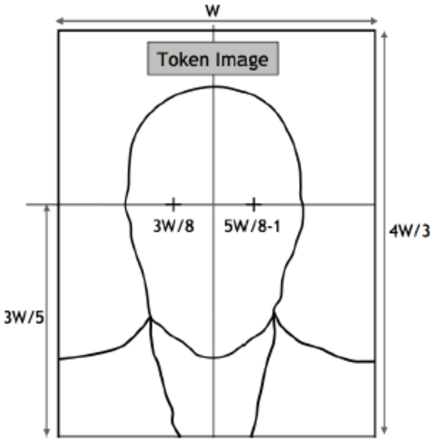
Operational backdrop

- Operators seek to collect face reference photos that will support high accuracy face recognition. Stored
  - in databases
  - on ID credentials

- Operators often require the reference image be a frontal portrait, conforming to requirements of an ISO standard, ISO/IEC 19794-5.
  - US and UK passports
  - DC driving licenses.

- Quality assessment is often manual (photographer, consular officer), more rarely automatic (with commercial software)



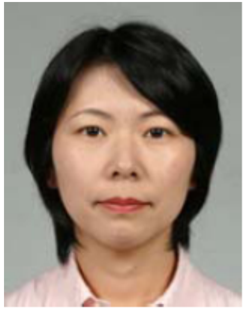


ISO/IEC 19794-5 Token Face Geometry, photometry, behavior are all regulated

Image dimensions, eye and head position are all parametric on W

Alternative standard views possible, in principle, but that ship sailed c. 2004.

Standards – And deviations from…



ISO Standard

Expression

Gaze

Too close

Pose Angle

…

NON-CONFORMANT EXAMPLES

- ISO's idea of "poor" images is better than any image contemplated in Janus.
- ISO aspires to collect reference samples that are pristine, for storage in authoritative databases.

Janus: Good, bad, wild, ugly, and lots beyond

Declining Quality → Declining Accuracy

WILD, PUBLIC SPACE

ISO[*] MUGSHOT[+] WEBCAM PRO-CAM[u] LEGACY-CAM[+] REFLECTED[x]

* http://webstore.ansi.org

\+ http://www.chicagonow.com/cta-tattler/2013/07/chicago-cops-use-face-recognition-software-to-nab-cta-mugger

u http://vis-www.cs.umass.edu/lfw/

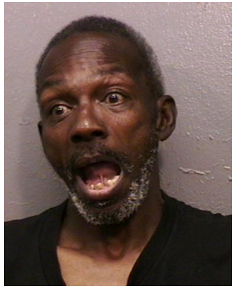
x http://io9.com/hidden-faces-can-be-found-by-zooming-into-hi-res-photos-1491607189

Part 2:
Face Image Quality Vectors

Face Image Quality Assessment: Standard #2

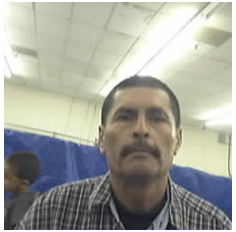
NIST

Vector Quality: Quantitative checks of subject and image properties



Component Image Quality Analysis

- Subject Behavior
  - Expression 70
  - Yaw 85
  - Pitch 80
  - Eyes-open 60
  - Glasses 98
  - Motion 97
- Camera + Environment
  - Illumination 34
  - Uniformity 68
  - Resolution 70
  - etc etc



Next steps

- Revise ISO/IEC 29794-5
    - Last update was 2010
    - Convert from technical report to standard
    - Add content
    - Align with new criteria in imminent ICAO Portrait Quality standard

- Timeline:
    - Initiate revision 2019-01
    - Push content: 2019-06
    - Substantially complete: 2020 late
    - Publication: 2021

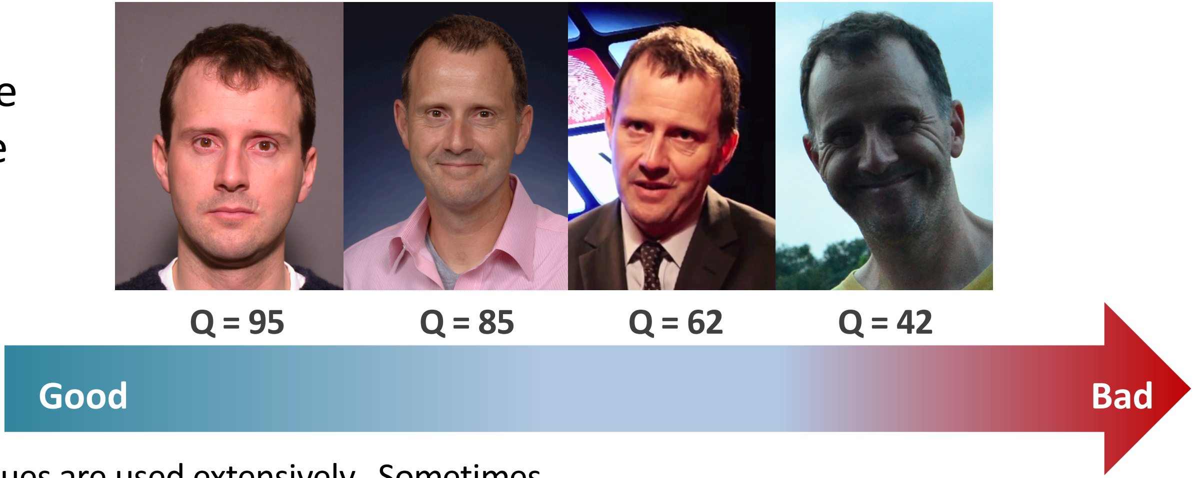
Part 2:
Face Image Quality Vectors

Operational need:

- Populate authoritative face repositories with photographs that will support high accuracy face recognition.

- The reference photo is widely specified as a frontal portrait, conforming to requirements of an ISO standard, ISO/IEC 19794-5.

Scalar Quality: Single value
Represents utility of image
to a recognition engine



In fingerprint operations, quality values are used extensively. Sometimes attending operators are paid by based on quality statistics.

Image quality values use-case 1 of 3: Capture Review

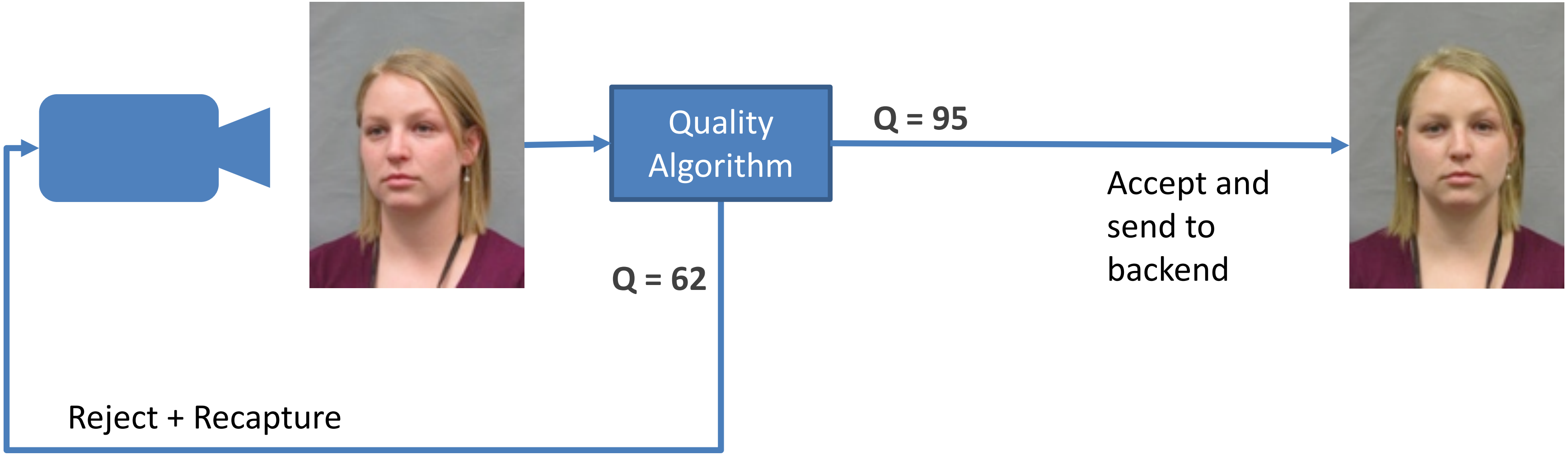


Image quality values use-case 2 of 3: Best sample selection

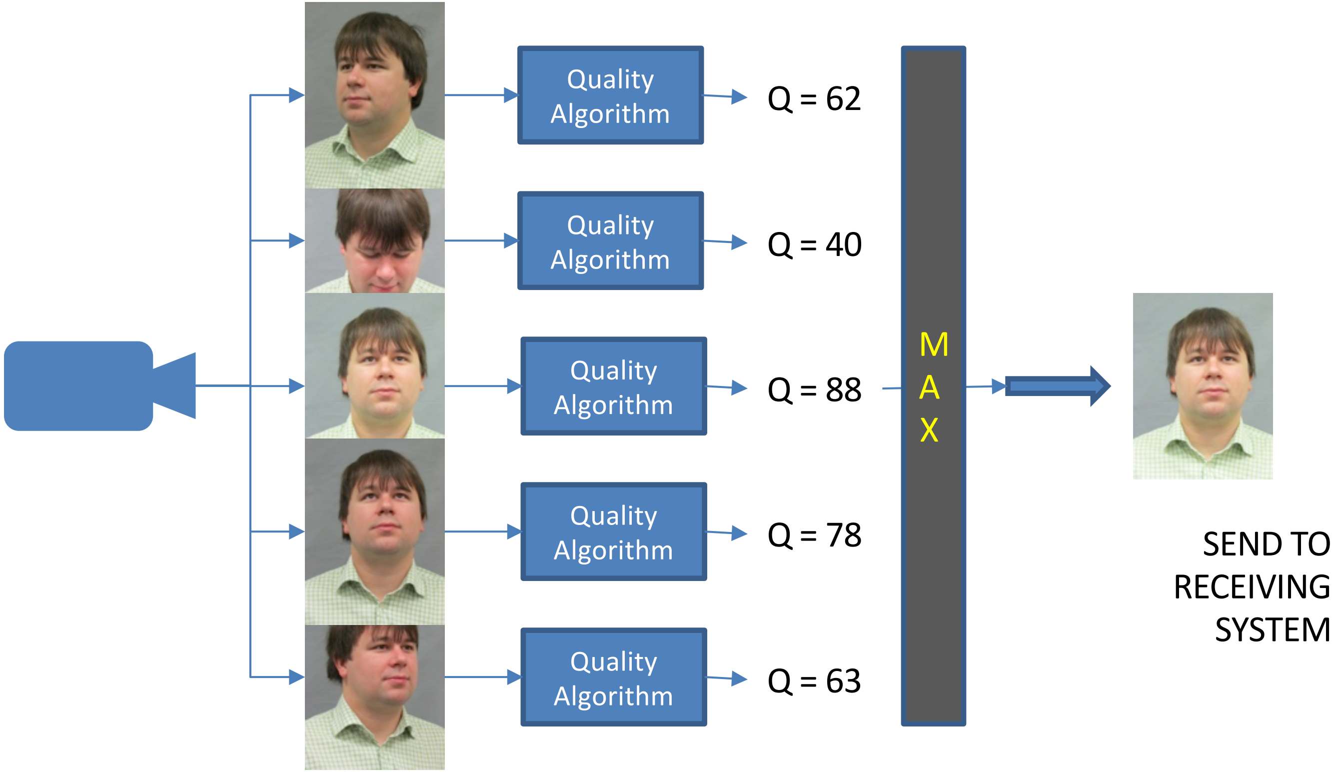


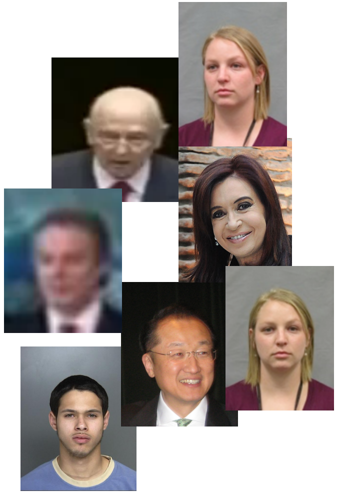
Image quality values use-case 3 of 3: Survey

Identify Variation Across:

- Locations
- Populations
- Sites
- Camera types

And with time

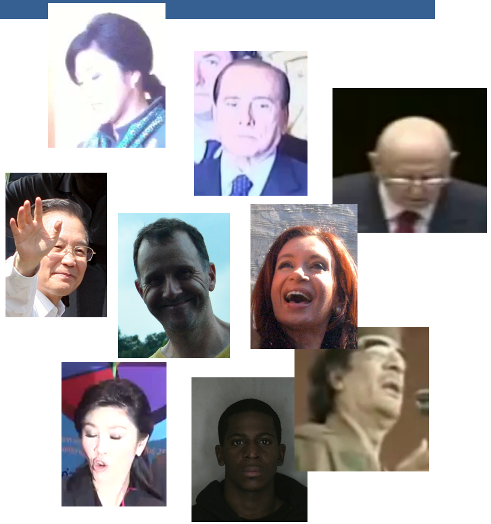
- Trends
- Seasonal
- Diurnal
- ...



Collection at Airport A
or
During Time Period A

Aggregate Q = 84

>



Collection at Airport B
or
During Time Period B

Aggregate Q = 53

Image Quality Vectors: Use case 2 of 2: Fault Detection

Identify prevalence of image problems across:

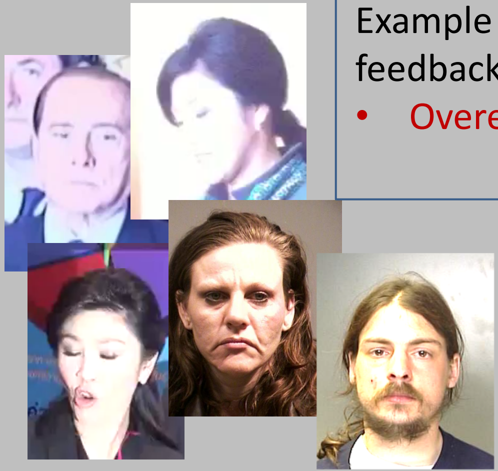
- Locations
- Populations
- Sites
- Camera types

And with time

- Trends
- Seasonal
- Diurnal
- …

Example 1: Specific feedback to site

- Glasses



Example 2: Specific feedback to site

- Overexposure

How can quality predict recognition success?

» A quality algorithm F operating on an image X_1 produces value

- $Q = F(X_1)$ [1]

» Face recognition algorithms compare samples to yield (genuine) scores

- $S = V(X_1, X_2)$ [2]

» Quality algorithms **shall** predict S from X_1 alone.

» Operating under the assumption that X_2 would be a canonical portrait image i.e. a pristine image of the same subject

- $Q \sim V(X_1, X_{PORTRAIT})$ [3]

- Respects the ISO/ICAO specification as the gold standard for AFR.
- The light grey text indicates that quality assessment must be done “blind”, targeting a hidden or virtual portrait image
  - cf. blind PSNR in image or video fidelity

Problem? Scores depend on *two* images: Individual / joint influence of degrading factors

Covariates where difference matters

FRR ~ F(Q(X) - Q(Y))

- Facial expression
- Non-frontal pose (sometimes)

- Beards on/off; Cosmetics on/off
- Eye-glasses (rims)

Broadly: Subject behaviors influential on facial properties

Covariates where the worse of two values matters

FRR ~ F(min(Q(X), Q(Y))

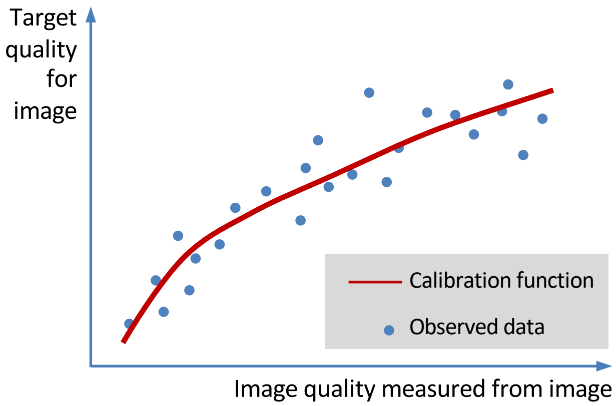
- Compression
- Blur
- Saturation
- Contrast, number of grey levels, entropy
- Occlusion (e.g. sunglasses)

Broadly: Image properties, photometric properties

Performance of an image quality algorithm

» What: NIST planning an open competition
- Expose poor algorithms i.e. those that produce random numbers
- Find best algorithms i.e. predictors of recognition accuracy
- Calibrate algorithms
  - Make Q scores meaningfully interpretable.
  - Support threshold decisions

» How:
- Run quality algorithms on LARGE sets of USG data
- Run recognition algorithms too
- Require quality algorithms to target recognition outcomes i.e. prediction

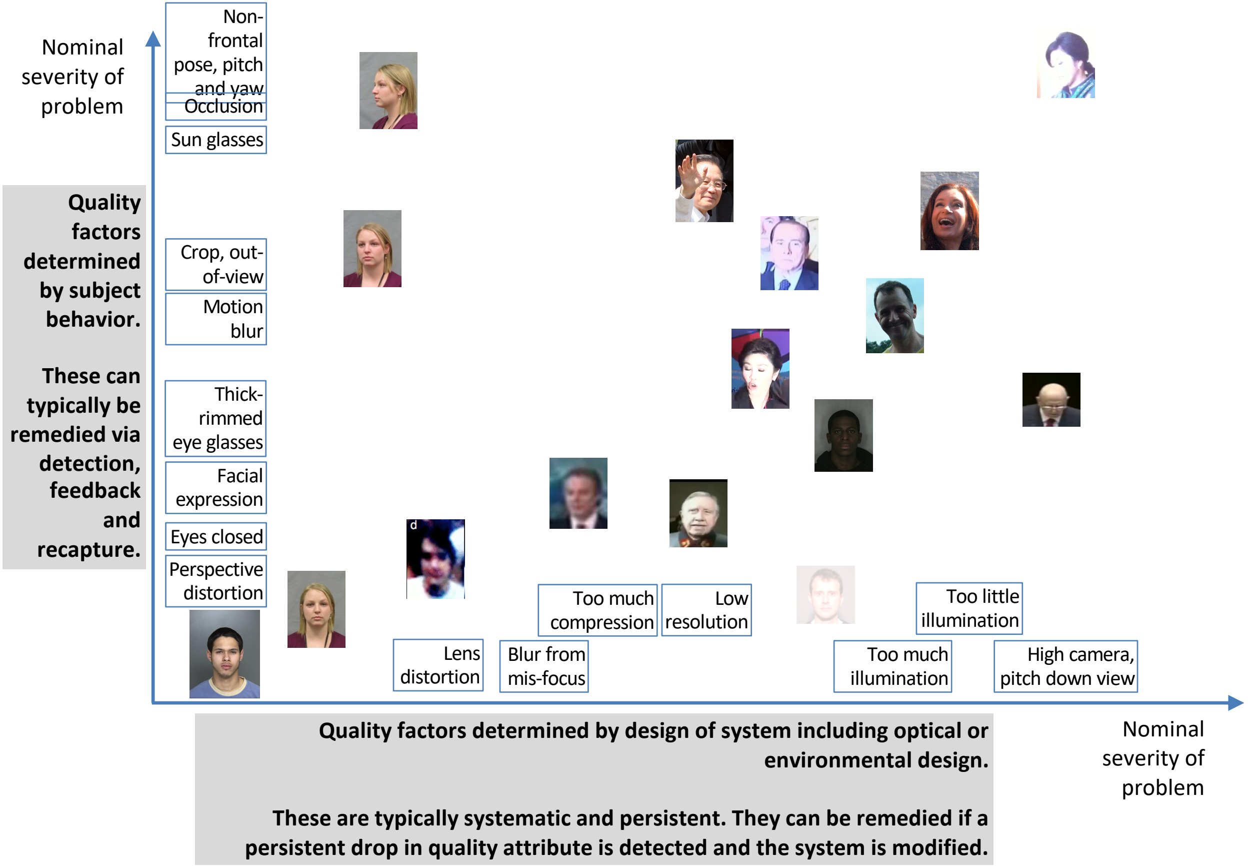


» When:
- Summer 2018

» Pre-requisites
- Review and consensus of quality value standard
- Representative large image datasets delivered to NIST

Now active: https://www.nist.gov/programs-projects/face-recognition-quality-assessment

Subject- vs. Imaging-specific problems



Part 3: Next Generation Capture Device Standard

Next generation capture drivers + capabilities

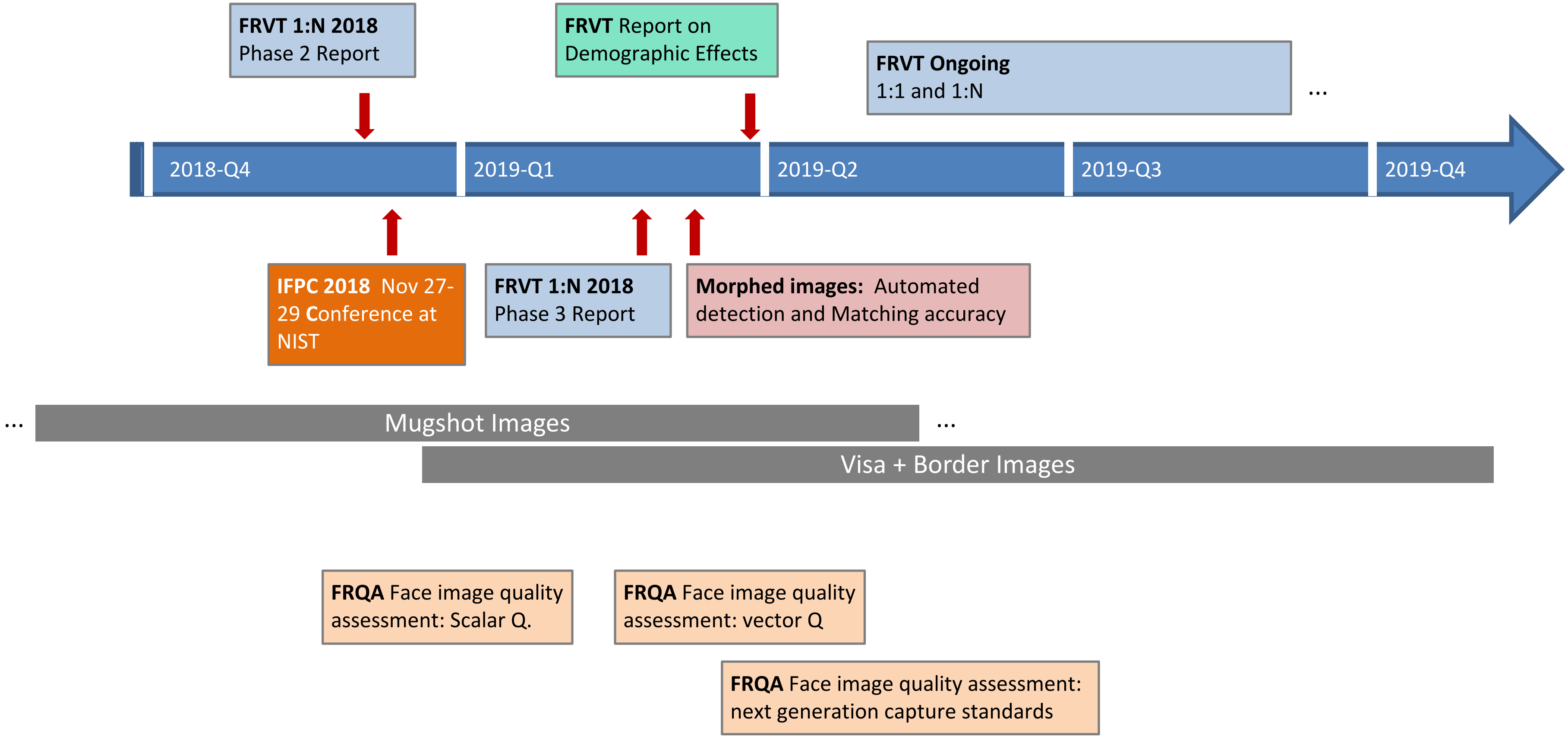
Background

- Increased reliance on face recognition
- Increased use globally
- Unlike fingerprint, iris, most face cameras are "dumb", unaware of the face itself
- Many photos deviate from ISO/ICAO
  - Subject appearance
  - Poor imaging
- Better recognition algorithms
  - But fail with pose, resolution, demographics
- Human "forensic" adjudication errors
- New opportunities for image manipulation

Capabilities

- Build auto-capture loop into camera
  - Face detector
  - Pose estimator
  - Correct exposure
  - Frontal views
- Collect high resolution for
  - Forensics adjudication
  - Morph Attack Detection
  - Presentation Attack Detection
- Collect non-frontal in some applications
- Prepare lower resolution for auto FR
  - 640 x 480 remains gold standard
- 3D, for accuracy, attack detection
- New compression
- Digital signatures to protect integrity

Face Recognition Activities at NIST



Thanks

pgrother@nist.gov
frvt@nist.gov

IARPA / NIST
Face Recognition Prize Challenge

F I V E

Face In Video Evaluation (FIVE)

FRVT
FACE RECOGNITION VENDOR TEST